Tugas Besar II Pengenalan Komputasi  
Asesmen Perusahaan Penyedia Jasa Analisa Limbah

Syarifa Khairunnisa  
Faculty of Industrial Technology  
Bandung Institute of TechnologyBekasi, Indonesia  
syarifakhrnsa18@gmail.com

Haryo Pandu Winoto, S.T.,M.Sc  
Faculty of Industrial Technology   
Bandung Institute of TechnologyBandung, Indonesia  
haryowinoto@office.itb.ac.id

*Abstract*—Pada tugas besar 2 ini penulis melakukan analisis terhadap dua macam perusahan penyedia jasa analisa limbah. Setelah analisis penulis dapat mengetahui statistik dari data perusahaan tersebut serta pola dan korelasi antar variabelnya. Hasil analisis ditujukan untuk memberikan saran kepada atasan penulis sebaiknaya memilih perusahaan mana yang paling baik untuk mendapatkan kontrak analisa limbah. Analisis deilakukan menggunakan bahasa pemrograman Python dengan bantuan modul Pandas, Matplotlib, dan Seaborn.

Keywords—Analisa limbah, statistik, python, pandas, data visualization.

# Introduction

Data merupakan komponen utama dari sistem informasi perusahaan karena semua informasi untuk pengambilan keputusan berasal dari data. Untuk mengolah data menjadi bentuk yang lebih bermanfaat dibutuhkan analisis yang tajam. Analisis data merupakan metode yang digunakan untuk mengetahui bagaimana menggambarkan data, hubungan data, semantik data dan batasan data yang ada pada suatu sisitem informasi. (Doro E, 2009)

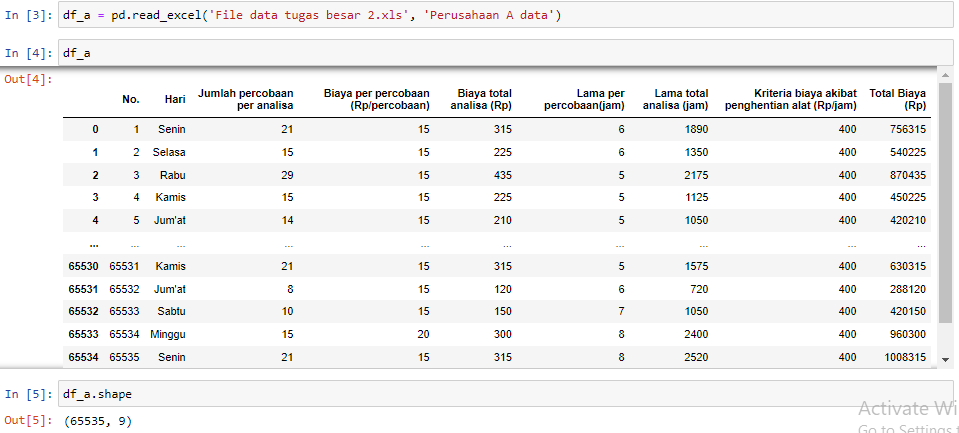
Pada zaman sekarang ini, analisis data adalah *skillset* penting yang sangat berguna dan banyak aplikasinya dalam berbagai bidang. Dengan analisis data, pengambilan keputusan akan berlandaskan pada fakta dan lebih terarah sehingga dapat memaksimalkan segala potensi dan meminimalisir resiko kerugian. Penelitian yang berlandaskan pada data ini menjadi salah satu pembelajaran yang membiasakan *data driven decision*.

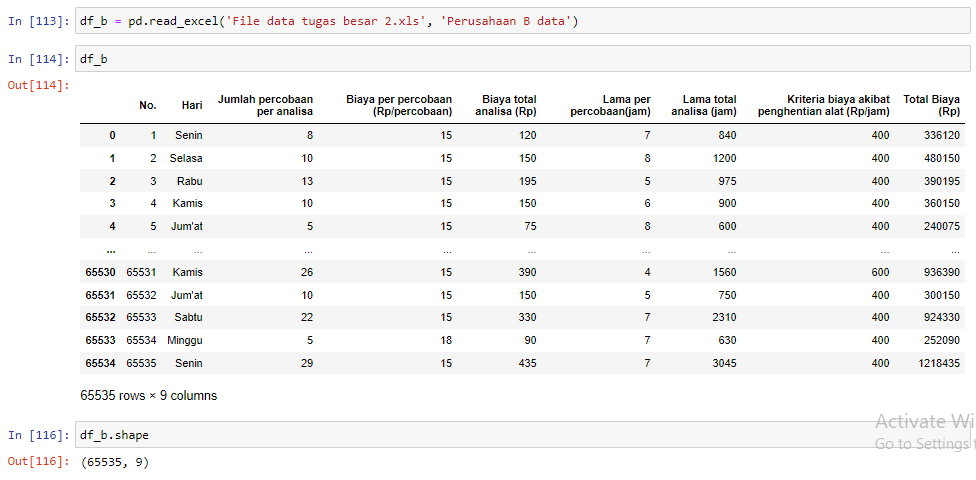
Sebagai tugas besar ke-2 mata kuliah Pengenalan Komputasi, penulis diberi tugas untuk melakukan analisis dari sebuah dataset. Dataset yang disediakan adalah data dua perusahaan penyedia jasa analisa limbah. Dari data ini penulis dapat mengetahui pola, korelasi, dan hal-hal lainnya yang dapat menjadi bahan pertimbangan untuk memilih salah satu dari kedua perusahaan tersebut.

# Data Description and Characteristics

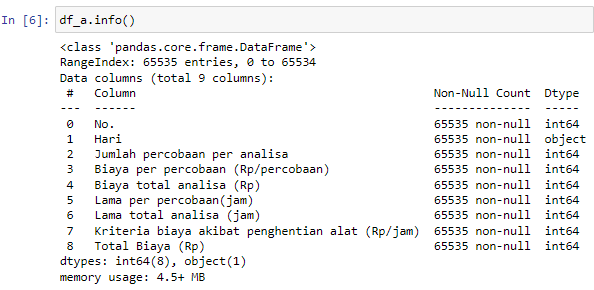
Data yang dianalisis adalah data excel dengan format .xls yang terdiri dari dua *sheets*. *Sheet* pertama adalah data perusahaan A dan *sheet* yang kedua adalah data perusahaan B. Pada kedua sheet dataset yang dianalisis terdapat sembilan atribut atau kolom dengan satu kolom kategorikal dan delapan kolom numerikal. Untuk analisis selanjutnya menggunakan bahasa python dan beberapa modul tambahan, *sheet* pertama di*import* sebagai “df\_a” dan *sheet* kedua di*import* sebagai “df\_b”.

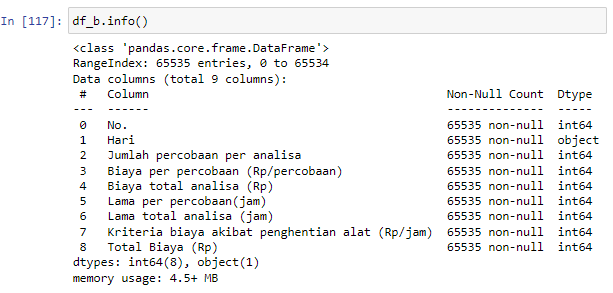
Berikut adalah gambaran besar dari kedua *sheet* data serta dimensi dari data yaitu 65.535 baris dan 9 kolom seperti yang terdapat pada gambar berikut



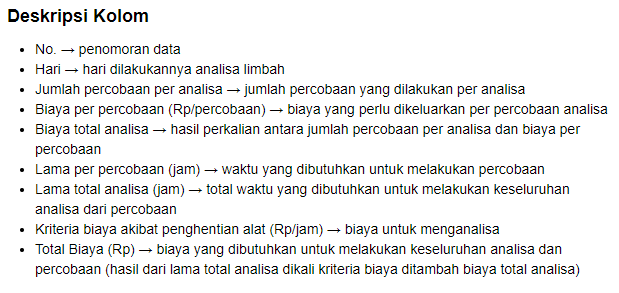


Ukuran data sekitar 13,237 kB dengan kolom-kolom sebagai berikut

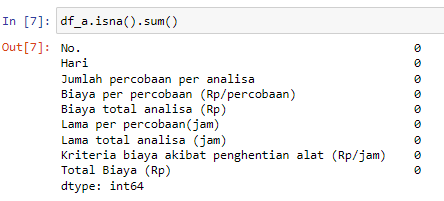


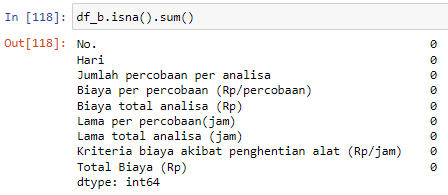


Dapat dilihat bahwa kedua data memiliki atribut atau kolom-kolom yang sama persis. Deskripsi dari tiap tiap kolomnya adalah sebagai berikut.



Sebelum melakukan analisis lebih lanjut, penulis melakukan pengecekan *missing value* terlebih dahulu untuk dilakukan *data preprocessing* karena keberadaan *missing value* dapat memengaruhi hasil akhir analisis (menyebabkan bias).

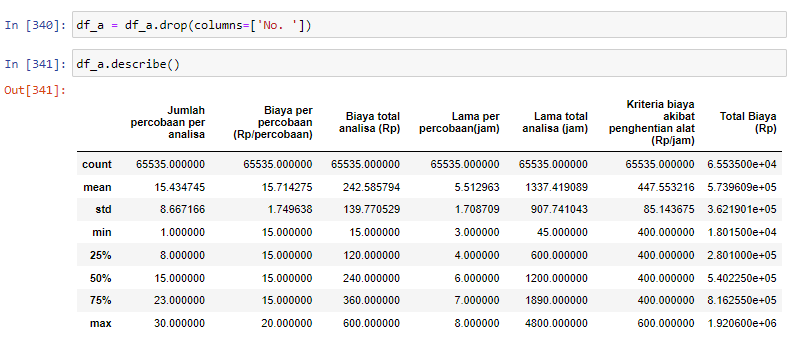


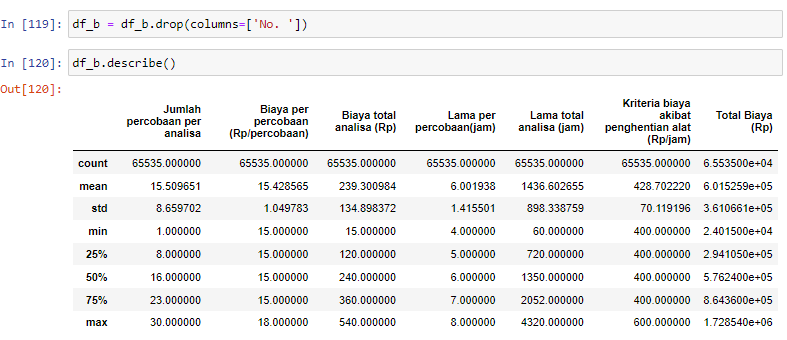


Berdasarkan hasil pengecekan, didapatkan bahwa pada kedua sheet data yang digunakan tidak terdapat *missing values* sehingga pada analisis kali ini tidak perlu dilakukan *handling missing values.*

# Data Statistics

## Statistik umum keseluruhan





Penulis melakukan *drop* pada kolom “No.” (membuang kolom tersebut) untuk memudahkan analisis selanjtunya karena kolom tersebut tidak mempunyai peran yang signifikan terhadap dataset. Berdasarkan atribut-atribut tersebut, penulis memutuskan untuk berfokus kepada variabel “Jumlah percobaan per Analisa”, “Lama total analisa (jam)”, dan “Total Biaya (Rp)” sebagai parameter penentu asesmen ini.

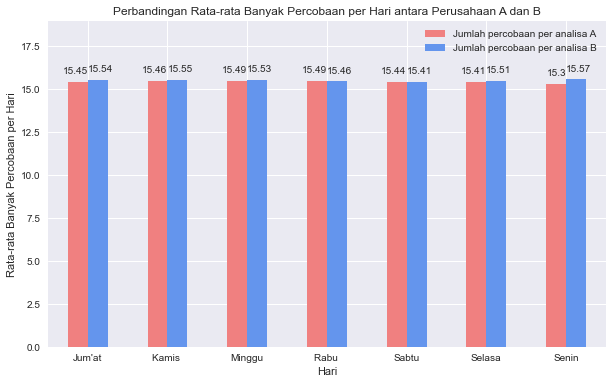
## Kesimpulan

Berdasarkan statistik umum keseluruhan data, rata-rata jumlah percobaan per analisa perusahaan A dan B tidak berbeda jauh yaitu 15,4 dan 15,5. Sedangkan lama total analisa perusahaan A adalah 1337,42 dan perusahaan B adalah 1436,6. Dapat dilihat bahwa rata-rata total biaya Perusahaan A lebih kecil dibandingkan Perusahaan B. Hal ini menunjukkan bahwa secara rata-rata keseluruhan, Perusahaan B memiliki biaya yang relatif lebih mahal dibandingkan dengan Perusahaan A, namun dari banyaknya percobaan dan lama analisa juga lebih tinggi. Detailnya akan dikaji pada bagian visualisasi data dengan pengelompokan perharinya.

# Data Visualization

## Perbandingan Kategori I

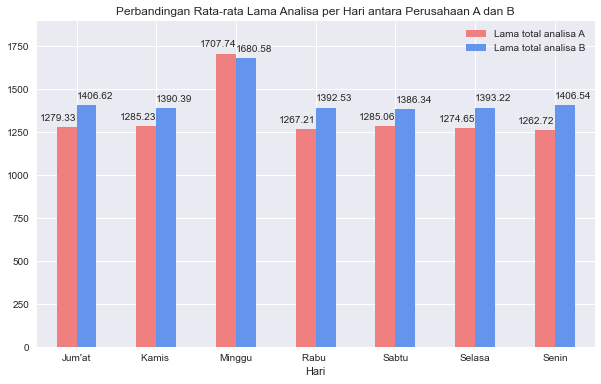
Penulis mengelompokan rata-rata banyaknya percobaan per hari kedalam tiap-tiap harinya dari masing-masing perusahaan kemudian menggabungkannya menjadi *dataframe* baru untuk dibandingkan dalam bentuk *bar chart*.



## Insight : Dapat dilihat bahwa perbedaan rata-rata banyak percobaan per hari antara Perusahaan A dan B tidak terlalu signifikan. Namun bila dilihat lebih jeli, pada hari Jum’at, Minggu, Selasa, dan Senin grafik Perusahaan B lebih tinggi dibanding perusahaan A. Hal ini menunjukkan secara garis besar, Perusahaan B lebih banyak melakukan percobaan dibandingkan dengan Perusahaan A.

## Perbandingan Kategori II

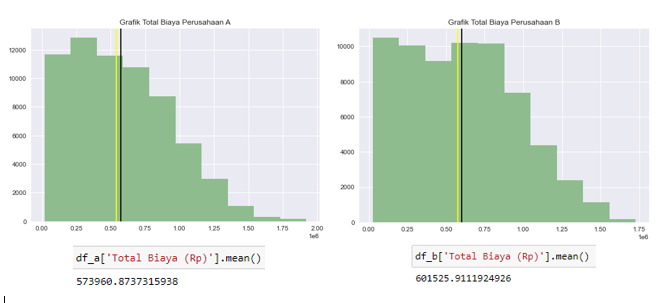
Penulis mengelompokan rata-rata lama analisa per hari kedalam tiap-tiap harinya dari masing-masing perusahaan kemudian menggabungkannya menjadi dataframe baru untuk dibandingkan dalam bentuk bar chart. Penulis mengasumsikan bahwa semakin lama durasi analisa maka semakin detail limbah tersebut dianalisa.



Insight: Dapat dilihat dilihat bahwa perbedaan lama total analisa di Perusahaan A dan B cukup signifikan dimana secara garis besarnya rata-rata analisa di Perusahaan B per harinya lebih lama dibandingkan Perusahaan A. Pada hari minggu, analisa di perusahaan A lebih lama namun besarnya tidak terlalu signifikan jika dibandingkan dengan hari-hari lainnya.

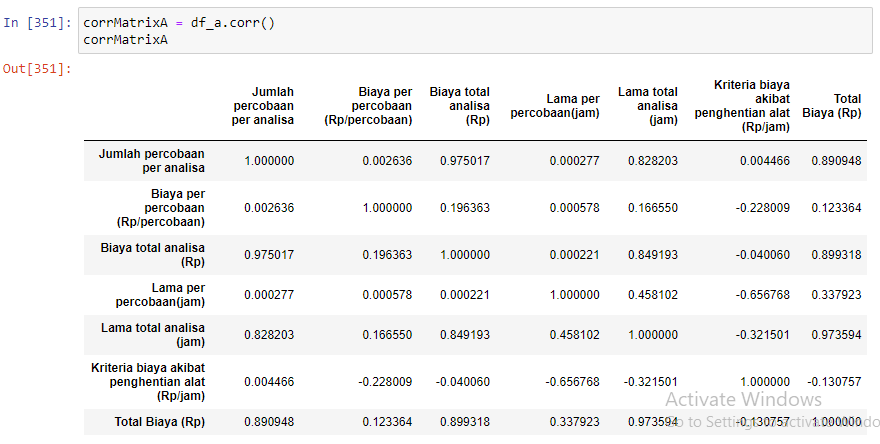
## Perbandingan Kategori III

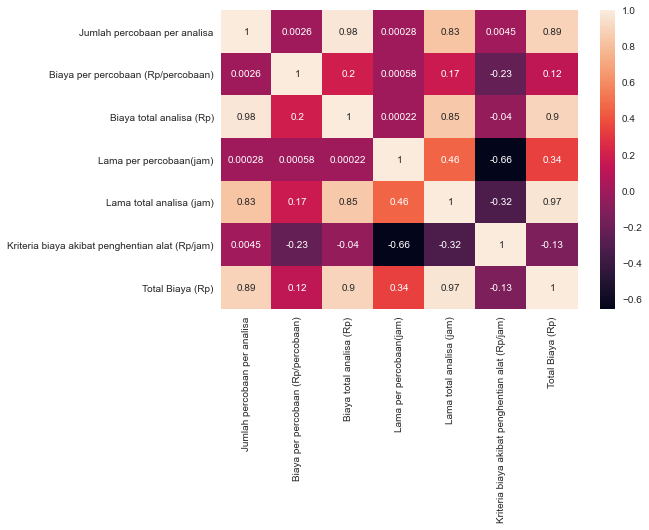
Penulis membandingkan histogram (persebaran) total biaya antara Perusahaan A dan B dengan garis kuning menunjukkan median dan garis hitam menunjukan mean (rata-rata) total biaya.



Insight: Persebaran total biaya di Perusahaan A relatif lebih rendah dibandingkan dengan Perusahaan B. Secara garis besar dapat dilihat bahwa median dan rata-rata berada dalam rentang yang sama, namun ketika dicari persisnya, rata-rata total Biaya Perusahaan A adalah Rp. 573.960,87 sedangkan total biaya Perusahaan B adalah Rp. 601.525,91

## Korelasi Keseluruhan Antar Variabel Perusahaan A

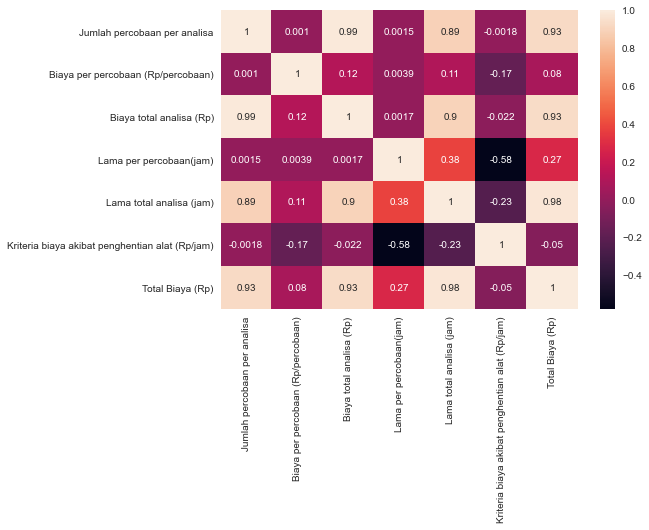




Insight: Pada Perusahaan A, variabel yang nilai korelasinya tinggi adalah “Jumlah percobaan per Analisa” dan “Biaya total analisa (Rp)”, “Jumlah percobaan per Analisa” dan “Lama total Analisa (jam)”, “Jumlah percobaan per Analisa” dan “Total Biaya (Rp)”, “Biaya total Analisa” dan “Lama total Analisa (jam)”, ‘Biaya total Analisa” dan Total Biaya (Rp)”, “Lama total Analisa (Jam)” dan “Total Biaya (Rp)”.

## Korelasi Antar Variabel Perusahaan B



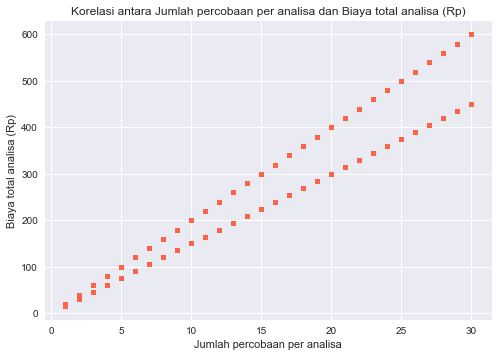


Insight: Pada Perusahaa B, variabel yang nilai korelasinya tinggi serupa dengan Perusahaan A, yaitu “Jumlah percobaan per Analisa” dan “Biaya total analisa (Rp)”, “Jumlah percobaan per Analisa” dan “Lama total Analisa (jam)”, “Jumlah percobaan per Analisa” dan “Total Biaya (Rp)”, “Biaya total Analisa” dan “Lama total Analisa (jam)”, ‘Biaya total Analisa” dan Total Biaya (Rp)”, “Lama total Analisa (Jam)” dan “Total Biaya (Rp)”.

# Correlations

## Korelasi antara Jumlah percobaan per analisa dan Biaya total analisa Perusahaan A





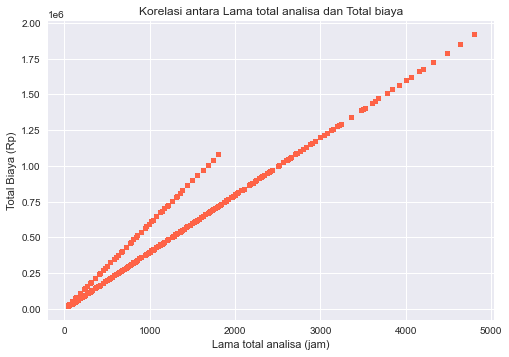
## Korelasi antara Lama total analisa dan Lama per percobaan Perusahaan A



## 

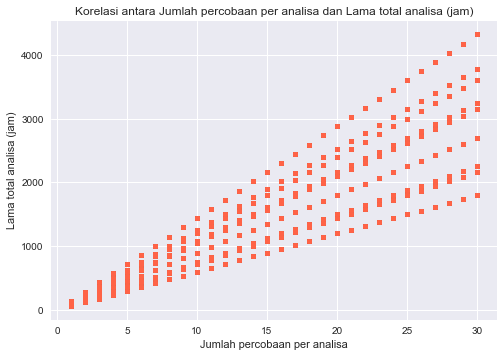
## Korelasi antara Lama total analisa dan Total Biaya Perusahaan A





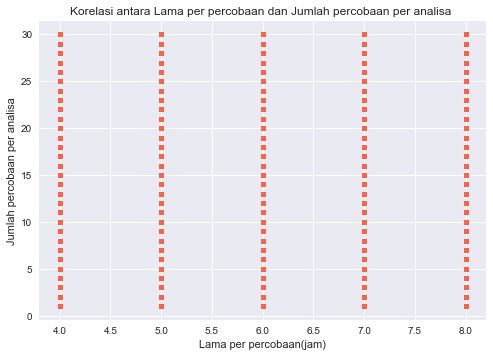
## Korelasi antara Jumlah percobaan per analisa dan Lama total analisa Perusahaan B





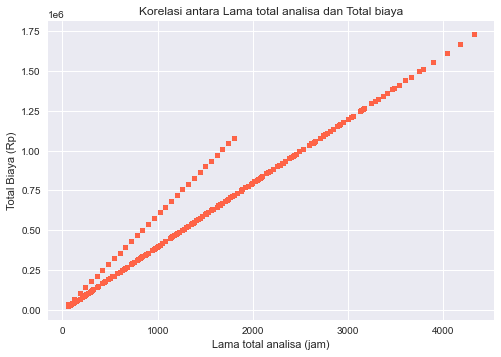
## Korelasi antara Lama per perccobaan dan Jumlah percobaan per analisa Perusahaan B





## Korelasi antara Lama total analisa dan Total Biaya perusahaan B

## 

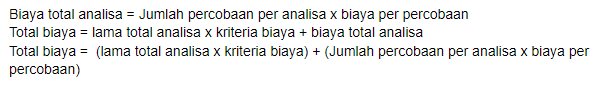


## Kesimpulan

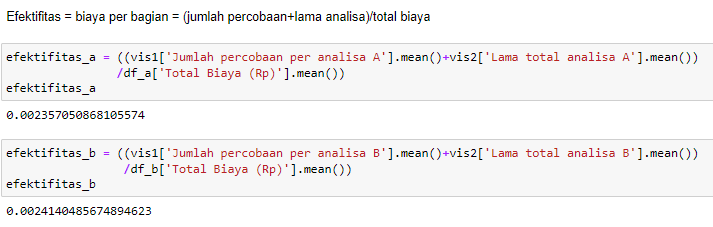
Berdasarkan scatter plot dari beberapa variabel diatas dapat dengan mudah diketahui mana varabel yang berkorelasi tinggi dan yang tidak. Desain seperti ini dapat menjadi kajian lebih lanjut dari *correlation matrix* keseluruhan.

# Efficiency Comparison

Variabel total biaya didapat dari hasil lama total analisa dikali kriteria biaya ditambahh biaya total analisa. Sedangkan biaya total analisa didapat dari hasil kali antara jumlah perdobaan per analisa dan biaya per percobaan. Skemanya dapat dilihat pada gambar berikut.



Penulis berfokus pada variabel lama total analisa, jumlah percobaan per analisa, dan total biaya. Maka secara tidak langsung dapat dirumuskan bahwa Total biaya = lama total analisa + jumlah percobaan per analisa. Untuk membandingkan mana yang lebih efisien, dilakukan perhitungan untuk mendapatkan hasil harga per bagian dengan rata-rata dari masing masing variabel seperti berikut.



Didapatkan bahwa harga per bagian dari Perusahaan A dan B hampir sama besarnya yaitu 0,002357 dan 0,002414.

# Conclusion

Berdasarkan analisis terhadap dataset yang telah dilakukan, secara rata-rata keseluruhan didapatkan hasil bahwa berdasarkan variabel Total biaya, perusahaan A relatif lebih murah dibandingkan dengan Perusahaan B. Namun hal tersebut berbanding lurus dengan kuantitas dan durasi jasa yang ditawarkan yaitu banyaknya percobaan dan lama analisa yang dilakukan. Perusahaan A lebih murah karena banyaknya percobaan dan lama analisanya lebih kecil dibandingkan dengan Perusahaan B.

*Ada kuantitas ada harga* merupakan gambaran secara garis besar, hal ini perlu dikaji lebih dalam lagi karena bisa jadi justru harga per bagian Perusahaan A lebih mahal dibandingkan Perusahaan B. Hasil perhitungan harga per bagian dari data menunjukkan bahwa keduanya tidak berbeda jauh.

Hasil yang dapat disimpulkan mengenai asesmen kedua perusahaan tersebut adalah, pemilihan perusahaan penyedia jasa analisa limbah dapat disesuaikan dengan apa yang dibutuhkan. Bila perusahaan kita menghasilkan limbah yang relatif banyak dan memiliki anggaran yang cukup, Perusahaan B dapat menjadi pilihan dengan pertimbangan bahwa kuantitas dan durasi analisa di perusahaan tersebut lebih tinggi sehingga limbah dapat dianalisa dengan baik. Bila perusahaan kita menghasilkan limbah yang tidak terlalu banyak, Perusahaan A dapat menjadi pilihan dengan pertimbangan kuantitas dan durasi yang lebih sedikit tersebut sudah cukup untuk menganalisa limbah yang tidak terlalu banyak.

# References

[1] Doro, Stevalin Betshani. 2009. *Analisis Data dengan Menggunakan ERD dan Model Konseptual Data Warehouse.* Fakultas Teknoloi Informasi Universitas Krinsten Maranatha. Bandung.